

Moritz Böhle

PHD STUDENT · MAX PLANCK INSTITUTE FOR INFORMATICS

✉ moritzboehle@me.com | 📱 moboehle | 🌐 moritz-boehle

Research statement

My research is focused on interpretable, trustworthy, and responsible deep learning. In particular, over the course of my PhD, I investigated how to design *inherently interpretable* deep neural networks such as the CoDA and the B-cos Networks, which are optimised to inherently provide explanations that highlight important input features.

Education

Max Planck Institute for Informatics

PHD IN INTERPRETABILITY IN DEEP LEARNING, EXPECTED GRADUATION: 03/2024

Advisors: Prof. Dr. Bernt Schiele, Prof. Dr. Mario Fritz

Saarbrücken

03/2019 - present

Bernstein Center for Computational Neuroscience

MSC COMPUTATIONAL NEUROSCIENCE, GPA 1.0 (BEST POSSIBLE: 1.0)

Machine Learning · Models of Higher Brain Functions · Models of Neural Systems

Thesis: Noise Suppression and Speech Enhancement using Deep Learning. Grade: 1.0 (Best possible: 1.0)

Berlin

10/2016 - 01/2019

University of California, Santa Cruz

UC EDUCATION ABROAD PROGRAM, GPA 4.0 (BEST POSSIBLE: 4.0)

Programming in Java, C, C++ · Biophysics

Santa Cruz, CA, USA

10/2014 - 06/2015

Freie Universität Berlin

BSC PHYSICS, GPA 1.0 (BEST POSSIBLE: 1.0)

Linear Algebra · Analysis · Analytical Mechanics · Statistical Physics

Thesis: Evaluating different barrier-crossing theories using Langevin simulations. Grade: 1.0 (Best possible: 1.0)

Berlin

01/2012 - 08/2016

Research and Publications

(* EQUAL CONTRIBUTION)

- 2023 A. Parchami-Araghi*, **M. Böhle***, S. Rao*, B. Schiele. *Good Teachers Explain: Explanation-Enhanced Knowledge Distillation*. **Under submission.**
- 2023 S. Gairola, **M. Böhle**, F. Locatello, B. Schiele. *Investigating the Interpretability of Self-Supervised Visual Pretraining*. **Under submission.**
- 2023 **M. Böhle**, N. Singh, M. Fritz, B. Schiele. *B-cos Alignment for Inherently Interpretable CNNs and Vision Transformers*. arXiv:2306.10898. **Under submission.**
- 2023 S. Rao, **M. Böhle**, B. Schiele. *Better Understanding Differences in Attribution Methods via Systematic Evaluations*. arXiv:2303.11884. **Under submission.**
- 2023 S. Rao*, **M. Böhle***, A. Parchami-Araghi, B. Schiele. *Studying How to Efficiently and Effectively Guide Models with Explanations*. International Conference on Computer Vision (**ICCV**), 2023.
- 2023 A. Kukleva*, **M. Böhle***, B. Schiele, H. Kuehne, C. Rupprecht. *Temperature Schedules for self-supervised contrastive methods on long-tail data*. International Conference on Learning Representations (**ICLR**), 2023.
- 2022 **M. Böhle**, M. Fritz, B. Schiele. *B-cos Networks: Alignment is All We Need for Interpretability*. Conference on Computer Vision and Pattern Recognition (**CVPR**), 2022.
- 2022 S. Rao, **M. Böhle**, B. Schiele. *Towards Better Understanding Attribution Methods*. Conference on Computer Vision and Pattern Recognition (**CVPR**), 2022.
- 2022 **M. Böhle**, M. Fritz, B. Schiele. *Optimising for Interpretability: Convolutional Dynamic Alignment Networks*. IEEE Transactions on Pattern Analysis and Machine Intelligence (**TPAMI**), 2022.

- 2021 **M. Böhle**, M. Fritz, B. Schiele. *Convolutional Dynamic Alignment Networks for Interpretable Classifications*. Conference on Computer Vision and Pattern Recognition (**CVPR**), **oral**, 2021.
- 2019 **M. Böhle***, F. Eitel*, M. Weygandt, K. Ritter. *Layerwise Relevance Propagation for Explaining DNN Decisions in MRI-based Alzheimer's Disease Classification*. **Frontiers in Aging Neuroscience** 11 (2019): 194.

Professional Experience

- 05/2018 - 12/2018 **Deep Learning Research, Master's student**, Audatic GmbH, Berlin
- 10/2017 - 05/2018 **Software Development**, Digital Unit Volkswagen Financial Services AG, Berlin
- 01/2012 - 10/2012 **Software Development and Quality Management**, Tembit Software GmbH

Academic Activities

- 07/2023 - present **Co-supervision PhD student**, Collaboration with N. Pham, S. Rao, B. Schiele.
Concept-based explanations of Deep Neural Networks.
- 06/2023 - present **Co-supervision MSc student**, Collaboration with S. Arya, S. Rao, B. Schiele.
Increasing the interpretability of conventional Deep Neural Networks
- 11/2022 - present **Co-supervision PhD student**, Collaboration with S. Gairola, F. Locatello, B. Schiele.
Interpretable self-supervised, object-centric learning
- 01/2022 - present **Co-supervision MSc student**, Collaboration with A. Parchami-Araghi, S. Rao, B. Schiele.
Explanation-based Knowledge Distillation
- 01/2021 - present **Co-supervision PhD student**, Collaboration with S. Rao, B. Schiele.
Better understanding and utilising attribution-based model explanations
- 05/2021 - 10/2021 **Co-supervision BSc student**, Collaboration with N. Singh, D. Stutz, B. Schiele.
Exploring the relationship between robustness and interpretability
- 04/2020 - 03/2021 **Teaching Assistant**, Elements of Data Science and Artificial Intelligence, B. Schiele.
- 03/2019 - present **Reviewing activities**, IEEE PAMI, CVPR, ECCV, ICLR, NeurIPS, IEEE Trans. Inf. Forensics Secur., ICML
- 10/2015 - 09/2016 **Teaching Assistant**, Linear Algebra and Analysis, R. Klein

Invited Talks

- 03/2023 **Spotlight @ Explainability in Machine Learning Workshop, Tübingen.**,
Towards B-cos Deep Neural Networks as the new default
- 06/2022 **Spotlight @ XAI for Computer Vision (XAI4CV) workshop, CVPR.**,
B-cos Networks: Alignment is All We Need for Interpretability
- 12/2021 **Invited talk @ Massachusetts Institute of Technology (MIT).**,
B-cos Networks: Alignment is All We Need for Interpretability

Scholarships, Awards, Grants

- 10/2022 **Outstanding Reviewer**, ECCV 2022
- 01/2015 - 03/2019 **Full scholarship**, German Academic Scholarship Foundation (Studienstiftung des Deutschen Volkes)
- 10/2014 - 12/2014 **Sudy abroad scholarship**, DAAD PROMOS program
- 10/2014 **Travel grant**, Fulbright program

Extra-curricular activities

- 05/2021 - present **PhD student representative**, Computer Vision and Machine Learning department, MPI for Informatics
- 04/2017 - 01/2019 **MSc student representative**, Bernstein Center for Computational Neuroscience